BMC Genomics

CrossMark

Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies

Filippo Biscarini[1,8†], Nelson Nazzicari[1,2†], Marco Bink[3,10], Pere Arús[4], Maria José Aranzana[4], Ignazio Verde[5], Sabrina Micali[5], Thierry Pascal[6], Benedicte Quilot-Turion[6], Patrick Lambert[7], Cassia da Silva Linge[7], Igor Pacheco[7,9], Daniele Bassi[7], Alessandra Stella[1,8] and Laura Rossini[1,7*]

Abstract

Background: Highly polygenic traits such as fruit weight, sugar content and acidity strongly influence the agroeconomic value of peach varieties. Genomic Selection (GS) can accelerate peach yield and quality gain if predictions show higher levels of accuracy compared to phenotypic selection. The available IPSC 9K SNP array V1 allows standardized and highly reliable genotyping, preparing the ground for GS in peach.

Results: A repeatability model (multiple records per individual plant) for genome-enabled predictions in eleven European peach populations is presented. The analysis included 1147 individuals derived from both commercial and non-commercial peach or peach-related accessions. Considered traits were average fruit weight (FW), sugar content (SC) and titratable acidity (TA). Plants were genotyped with the *9K IPSC* array, grown in three countries (France, Italy, Spain) and phenotyped for 3–5 years. An analysis of imputation accuracy of missing genotypic data was conducted using the software Beagle, showing that two of the eleven populations were highly sensitive to increasing levels of missing data. The regression model produced, for each trait and each population, estimates of heritability (FW:0.35, SC:0.48, TA:0.53, on average) and repeatability (FW:0.56, SC:0.63, TA:0.62, on average). Predictive ability was estimated in a five-fold cross validation scheme within population as the correlation of true and predicted phenotypes. Results differed by populations and traits, but predictive abilities were in general high (FW:0.60, SC:0.72, TA:0.65, on average).

Conclusions: This study assessed the feasibility of Genomic Selection in peach for highly polygenic traits linked to yield and fruit quality. The accuracy of imputing missing genotypes was as high as 96%, and the genomic predictive ability was on average 0.65, but could be as high as 0.84 for fruit weight or 0.83 for titratable acidity. The estimated repeatability may prove very useful in the management of the typical long cycles involved in peach productions. All together, these results are very promising for the application of genomic selection to peach breeding programmes.

Keywords: Peach (*Prunus persica*), Genome-enabled predictions, Fruit weight, Sugar content, Titratable acidity, Genotype imputation, Repeatability model

Background

Peach (*Prunus persica* L. Batsch) has been bred and cultivated for more than 4 000 years [1] and is both an important crop and a model species for the Rosaceae family [2]. The total world peach production was 21.6 million tonnes in 2014, of which 18.5% (4 million tonnes) from

Europe [3]. Like most fruit trees, peach is a perennial crop. Because of their long juvenile phase, breeding perennial plants is a complex task that requires careful planning and precise economic evaluations [4]. In addition, varieties need to be tested in multiple locations over multiple years to assess their adaptation to the geographical environment and their production potential. The generation interval in current peach breeding programmes can be up to 5–7 years [5], and this limits the genetic gain potentially

*Correspondence: laura.rossini@unimi.it
†Equal contributors
[1]PTP Science Park, Via Einstein - Loc. Cascina Codazza, Lodi, Italy
[7]Università degli Studi di Milano - DiSAA, Via Celoria 2, Milano, Italy
Full list of author information is available at the end of the article

Biscarini *et al. BMC Genomics* (2017) 18:432

Page 2 of 15

achievable per unit of time. An additional complication is that several relevant phenotypes are typically measured late in life (e.g. fruit size, plant yield, maturation time), thereby increasing the length and costs (e.g. keeping selection candidates) of peach breeding programmes.

In species with long breeding cycles, genomic selection bears the potential of improving selection efficiency – through e.g. reduced generation intervals, thereby speeding up genetic progress [6]. This was the major motivation behind the swift uptake of genomic selection in dairy cattle [7]. The relative high economic value of cattle, helped dairy cattle breeders pioneer the use of genomics in agriculture [8]. The constant decrease of sequencing costs and the availability of SNP genotyping technologies for an ever increasing number of species [9–11], has expanded the interest for genomic selection in modern breeding programmes. The genome sequence of peach is available [12], and an updated version has been recently released [13]. SNP chips for *P. persica* have also been designed [14].

The availability of SNP data allows –given a reference population that is both genotyped and phenotyped– for genomic predictions of unobserved phenotypes and genetic values for relevant traits in selection candidates, which is an essential element for the application of genomic selection to breeding. Genomic predictions for a variety of traits have been successfully modelled in a wide range of plant species e.g. the forage crop alfalfa [15], sugar beet [16, 17], loblolly pine [18], eucalyptus [19], including some important fruit trees like apple [20] and pear [21]. Traits considered focussed initially on yield and fruit size, but interest is growing also for traits related to fruit quality and response to environmental conditions, for life cycle traits (longevity, disease resistance, adaptability etc.), and for multiple-trait selection [22].

In perennial plants, the long life cycles and multiple records over successive years call for the modelling of repeated records. Not only the genetic/breeding value of plants is relevant for selection, but also the possibility of predicting the future performance, for management purposes. Repeatability models have found widespread application in animal breeding [23], while their use in plant breeding has been limited (e.g. maize [24]; cashew [25]).

In this paper, a repeatability model for genome-enabled predictions in eleven European peach populations is presented, where repetition refers to measurements in multiple successive years. Traits considered were fruit weight, and sugar content and acidity, which are key traits related to the quality of the fruit. To our knowledge, this is the first time that genomic predictions for any traits are reported in *P. persica*, and the first application overall of a repeatability model to genomic predictions in plants. The heritability, repeatability and predictive ability for the three phenotypic traits in each peach population

are reported. In addition, the accuracy of missing genotypes imputation has been estimated, and statistical issues related to genome-enabled predictions have been discussed.

Methods
Plant material and genotypes
From research fields in Italy, France and Spain, 1 147 peach plants from 11 crosses were available: four crosses from Italy (459 plants), two crosses from France (250 plants), and five crosses from Spain (438 plants). Italian crosses came from orchards of the University of Milan and of the Fruit Tree Research Centre (CREA-FRU) in Roma; French crosses from orchards at INRA-Avignon; Spanish crosses from orchards at IRTA in Lleida. The crosses were: Bolero x Oro (BxO), Max x Rebus028 (MxR028), PI91459 (NJ Weeping) x Bounty (WxBy), and IF7310828 x (IF7310828 x Ferganensis) (PxF) from Italy; Bolinha x Bolinha (BoxBo) and (SD40 x Summergrand) x Zéphyr (BC2) from France; Big Top x Armking (BtxAk), Belbinette x Nectalady (BbxNl), Big Top x Nectacross (BtxNr), MB1.73 x Earlygold (T1E) and MB1.73 x MB1.73 (TxE) from Spain. The parental "SD40" originated from a cross between *P. persica* and *P. davidiana*.

All plants were genotyped with the peach IPSC 9K SNP array [14, 26], with an average call-rate of 96.7%. Of the initial 1 147 samples, 57 had a call-rate $\leq$ 0.90 and were discarded. Of the initial 8 144 SNP markers, 2 068 SNP that were monomorphic in all populations or had a call-rate $\leq$ 0.90 were removed from the dataset, leaving 6 076 SNP for the analysis. The residual average missing rate was 3.19% (1.17% in BbxNl; 5.79% in TxE). A summary of plant populations (i.e. progenies) and genotype data can be found in Table 1.

Phenotypic data
Fruit weight, sugar content and acidity measurements were available for the 1 090 peach trees left after editing for call-rate from the 11 *P. persica* populations. Fruit weight (FW) in grams was measured as the average weight of 10 random peaches sampled from each tree. Sugar content (SC) and titratable acidity (TA) were measured, respectively, as average Brix degrees (soluble solid content) and meq/100 ml in the juice of at least five ripe fruits. FW was available for all 11 peach crosses; SC and TA only for 9 crosses (all except BoxBo and WxBy).

In most cases, phenotypic records for multiple years were available (most commonly two or three years; only one for FW in BoxBo, as many as 6 for FW in PxF), collected between 1995 and 2013. The number of records spanned from 16 (TA, TxE, 2011) to 127 (FW, BC2, 2005). Figure 1 shows the boxplots of the phenotypic distributions per peach population and year of measurement. All phenotypes were approximately normally distributed.

Biscarini *et al. BMC Genomics* (2017) 18:432

Page 3 of 15

Table 1 Distribution of plants per cross and summary of SNP-genotype data

Cross	Code	Origin	Samples	CR>0.9	MR	MAF=0	Mean MAF
Belbinette x Nectalady	BbxNl	Spain	102	102	0.01	48.90%	28.30%
Big Top x Armking	BtxAk	Spain	77	74	0.01	45.50%	44.10%
Big Top x Nectacross	BtxNr	Spain	51	43	0.03	49.00%	47.60%
Bolinha x Bolinha	BoxBo	France	115	111	0.03	42.20%	40.40%
Bolero x OroA	BxO	Italy	129	126	0.02	37.50%	34.60%
IF7310828 x (IF7310828 x Ferganensis)	PxF	Italy	130	130	0.04	22.10%	22.60%
Max10 x Rebus	MxR028	Italy	73	72	0.04	49.50%	32.70%
MB1.73 x Earlygold	T1E	Spain	148	124	0.04	9.40%	21.10%
MB1.73 x MB1.73	TxE	Spain	60	54	0.06	30.40%	27.50%
(SD40 x Summergrand) x Zéphyr	BC2	France	135	131	0.04	19.00%	24.00%
NJ Weeping x Bounty	WxBy	Italy	127	123	0.03	40.50%	36.20%
Entire dataset			1147	1090	0.03	2.20%	26.40%

CR > 0.9: call-rate > 0.90; *MR*: residual missing rate after editing on the whole dataset; *MAF* = 0: monomorphic SNP

Descriptive statistics on the measured phenotypes per cross and year are reported in Additional file 1.

Imputation of missing genotypes
After genotyping and editing for call-rate, residual missing genotypes were imputed using the localized haplotype clustering imputation (LHCI) method implemented in the software "Beagle" [27]. Originally developed for human genetics, LHCI has since found wide application also in animal and plant genetics (e.g. [11, 17, 20, 28]). Imputation was carried out in each cross separately, to avoid potential problems due to population heterogeneity.

The accuracy of imputation was measured. For each cross, a subset with no missing data was extracted from the total dataset, and increasing proportions of missing genotypes were then artificially introduced: 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 and 30%. Missing genotypes were imputed using the LHCI method in Beagle. For each missing rate and cross, the imputation was repeated 10 times, each time resampling randomly the genotypes to be set to missing. The average proportion of correctly imputed genotypes over the 10 replicates for each missing rate was then used to estimate an empirical curve of the imputation accuracy in each peach cross.

Assessment of the population structure
The population structure in the analysed peach population was assessed based on the kinship among all crosses. From imputed SNP genotypes, marker-based genomic relationships were estimated à la Astle & Balding [29]:

$$\mathbf{G} = \frac{1}{L} \sum_{l=1}^{L} \frac{\left(Z_{.,l}\right)\left(Z_{.,l}\right)'}{4p_l\left(1 - p_l\right)} \tag{1}$$

where L is the number of marker loci, $Z_{.,l}$ is the l_{th} column of the matrix of marker genotypes corrected by allele frequencies, and p_l is the allele frequency at locus l.
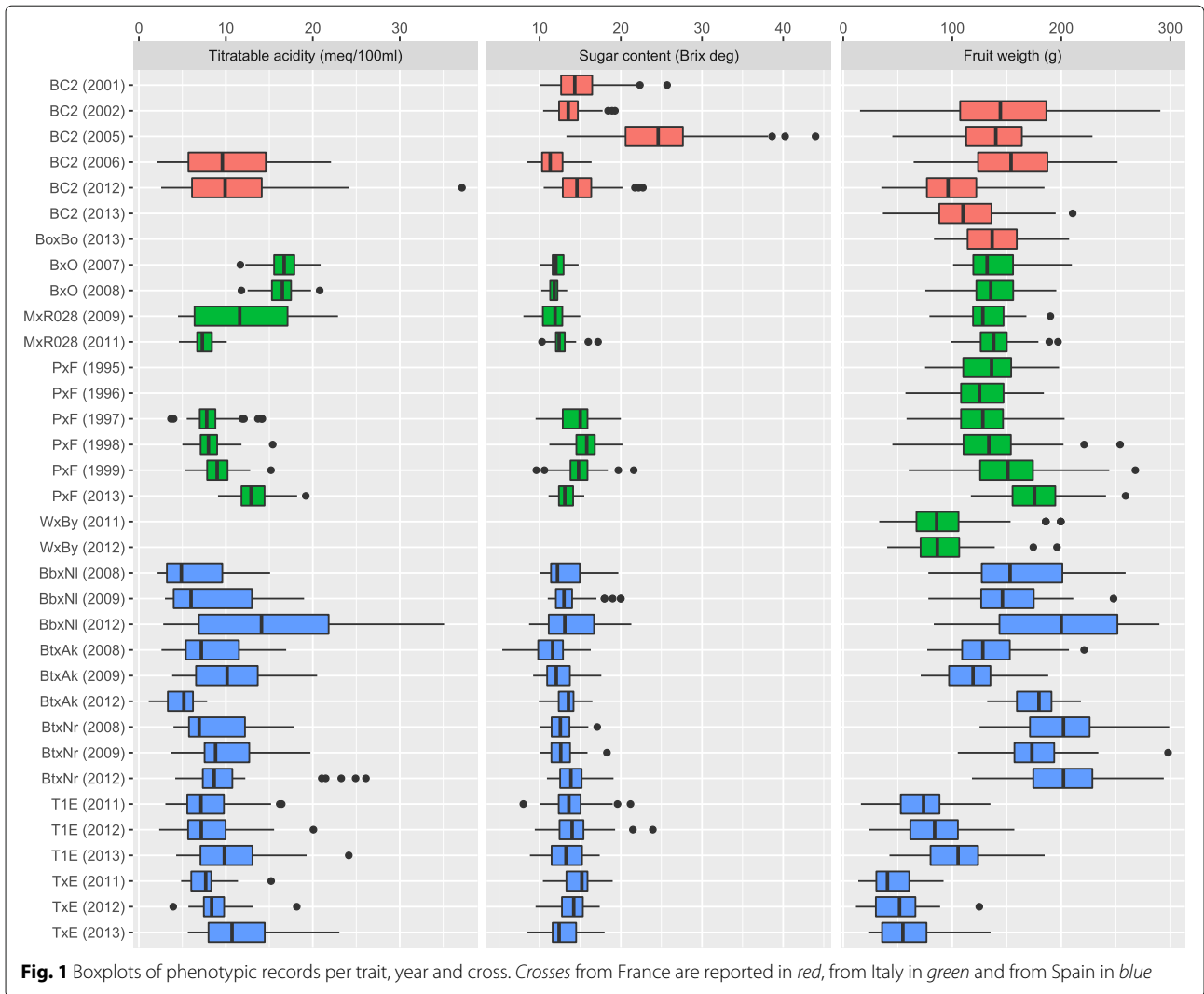
From the kinship matrix in Eq. 1, the Neighbor-Joining (NJ) tree [30] of the 11 peach crosses was constructed.

Repeatability model for genome-enabled predictions
For the prediction of fruit size, sugar content and acidity based on SNP genotypes, a GBLUP (Genomic Best Linear Unbiased Predictions) approach was used [31]. GBLUP was run for each peach population separately, and SNP with within-population MAF < 1% were removed before the analysis. Since multiple measurements for the same trait were recorded in successive years on individual peach trees, a repeatability model was used to fit systematic, additive genetic and permanent environment effects [23]. In matrix notation, the model had the following form:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e} \tag{2}$$

where $\mathbf{y}$ is the vector of (repeated) observations for each of the three traits; $\mathbf{b}$ is the vector of fixed effects: the overall mean and the year of measurement (categorical); $\mathbf{a}$ is the vector of additive genetic effects; $\mathbf{pe}$ is the vector of permanent environment effects; $\mathbf{e}$ is the vector of residual effects; $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{W}$ are incidence matrices which relate records in $\mathbf{y}$ to fixed, additive genetic and permanent environment effects, respectively. Residuals and permanent environment effects are assumed to be independent and normally distributed, with mean zero and variances $\mathbf{I}\sigma_e^2$ and $\mathbf{I}\sigma_{pe}^2$. The additive genetic effects are also assumed to follow a normal distribution, and have mean 0 and variance $\mathbf{G}\sigma_a^2$, where $\mathbf{G}$ is the matrix of genomic relationships –within cross– calculated as in Eq. 1. From

Biscarini *et al. BMC Genomics* (2017) 18:432

Page 4 of 15

Fig. 1 Boxplots of phenotypic records per trait, year and cross. *Crosses* from France are reported in *red*, from Italy in *green* and from Spain in *blue*

variance components, the narrow sense heritability (h^2) and repeatability (R) were derived:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2} \tag{3}$$

$$R = \frac{\sigma_a^2 + \sigma_{pe}^2}{\sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2} \tag{4}$$

For specific combinations of peach cross and trait the sample size was rather small in terms of number of records and, especially, number of unique individuals (see Additional file 1). This, at times, made matrices singular and non-invertible, leading to numerical problems with the estimation of parameters. Therefore, the model in Eq. 2 was fit using either Restricted Maximum Likelihood (REML: [32]) or an MCMC approach [33].

Estimation of the predictive ability

For each trait and cross, the predictive ability (*PA*) of model 2 was assessed through a 5-fold cross-validation.

Plant records were randomly partitioned into 5 subsets of approximately equal size (from $\sim$ 12 records for TA in TxE, to $\sim$ 120 records for FW in PxF). In turn, the records in one subset were set to missing and predicted using the model trained in the remaining four subsets, until all subsets were once used as validation set and every observation was used both to train and validate the model. The 5-fold cross-validation was repeated 100 times, each time resampling different subsets, eventually yielding 500 replicates of the analysis (per peach cross, per trait). In each replicate, h^2, R and the predictive ability were calculated.

PA was calculated as the correlation between observed and predicted phenotype, $r(\hat{y}, y)$, in the validation set. Predicted observations were obtained by summing up effects from the model (Eq. 2): $\hat{y} = \mu + y\hat{e}ar + \hat{a} + \hat{pe}$. Estimates of h^2, R and $r(\hat{y}, y)$ were averaged over the 500 replicates to obtain robust estimates of the central tendency and variability of the genetic parameters for fruit size and quality and of the accuracy of genomic predictions.

Biscarini *et al. BMC Genomics* (2017) 18:432

Page 5 of 15

Software used

Imputation of missing genotypes was performed using the *Beagle* software [27]. Variance components were estimated with a restricted maximum likelihood approach using the *Asreml* software [34] or with a MCMC approach using the BGLR R package [35]. Data manipulation, the parsing of results and plots were done using the *R* software [36].

Results

SNP genotypes and imputation accuracy

After imputation of residual missing genotypes, the proportion of monomorphic SNP ranged from 9.4% in T1E to 49.5% in MxR028, with an average of 35.8% over all crosses. After monomorphic SNPs were removed, the minor allele frequency (*MAF*) ranged from 21.20% (T1E) to 47.60% (BtxNr), with an average of 26.4% over all samples. Details on monomorphic markers frequency and MAF are reported in Table 1.
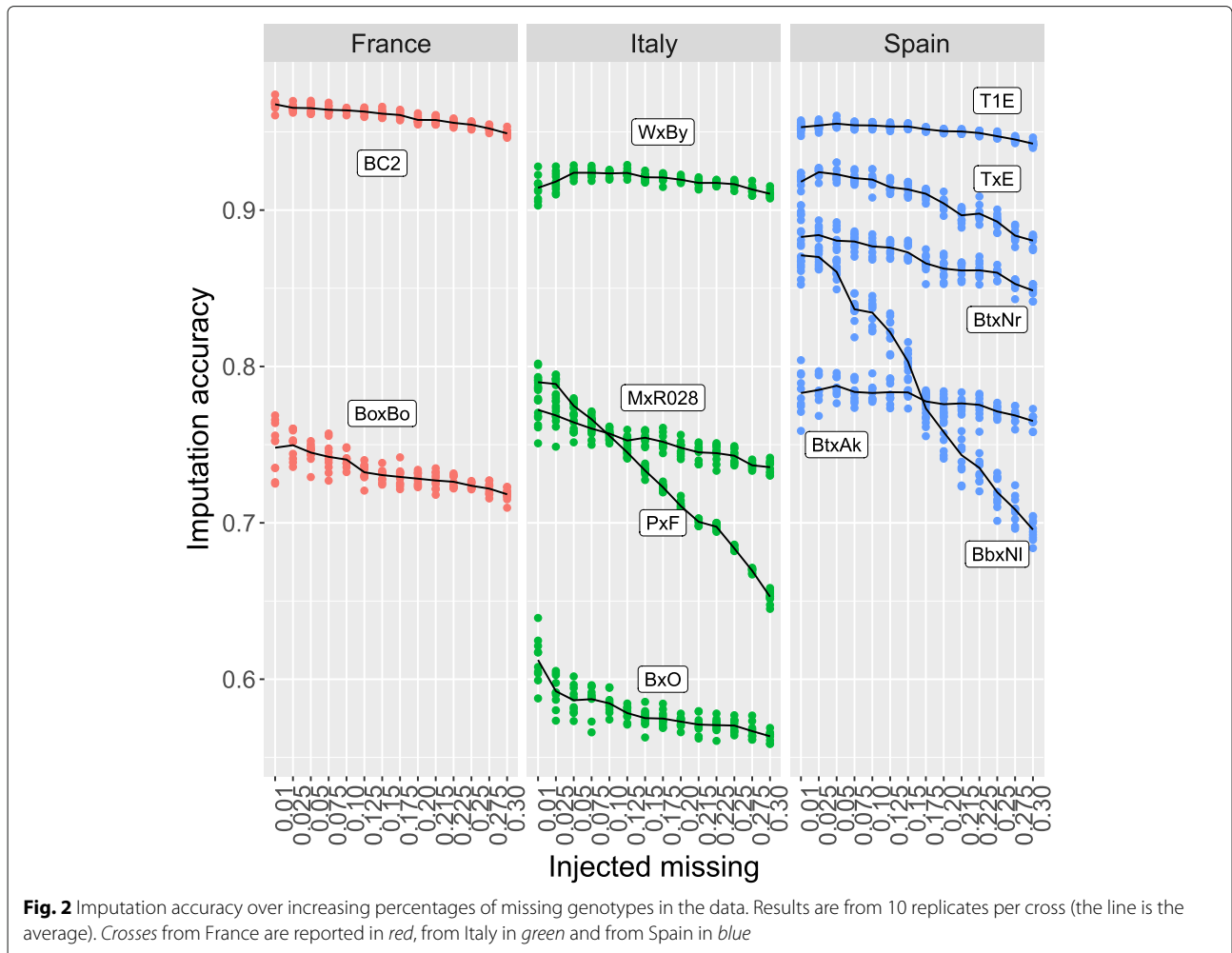
The imputation accuracy was measured per cross as the ratio of correctly imputed genotypes over the total number of artificially introduced missing genotypes, for

increasing missing rates (from 1 to 30%). Results from 10 repetitions (per cross, per missing rate) are reported in Fig. 2; the interpolating lines are the average accuracies. Standard deviations and further details can be found in Additional file 2.

The average imputation accuracy over all crosses and missing rates was 0.82, and varied from 0.96 (BC2) to 0.58 (BxO). Imputation accuracy was typically higher with low proportions of missing genotypes in the data: 0.97 in BC2 and 0.95 in T1E at 1% missing rate. The lowest imputation accuracy (0.564) was found for BxO at 30% missing genotypes. Most crosses showed a generally flat response to increasing missing rates, with imputation still performing well even with 20% or more missing genotypes. Exceptions were PxF and BbxNl, whose imputation accuracy dropped by 17 and 20 percentage points, respectively, between 1 and 30% missing rates.

Kinship matrix and neighbor-joining tree

From the multidimensional scaling of the reciprocal of the kinship matrix in Eq. 1 ($1 - G$), principal coordinates and

Fig. 2 Imputation accuracy over increasing percentages of missing genotypes in the data. Results are from 10 replicates per cross (the line is the average). *Crosses* from France are reported in *red*, from Italy in *green* and from Spain in *blue*

corresponding eigenvalues were obtained. The two first dimensions accounted for 27.5% of the genetic variability (50.3% the first five), and are plotted in Fig. 3: a strong population structure is apparent, with specific peach crosses clearly clustering together. BoxBo and WxBy clustered separately from other crosses: BoxBo formed a very compact cluster, while the WxBy cluster is more spread out. Five crosses (BbxNl, MxR028, BC2, BtxAk, BtxNr) formed an entangled cluster and higher-order dimensions are required to visually separate them. The highly structured nature of the dataset is confirmed by the Neighbor-Joining (NJ) tree [30] obtained from the across-population kinships and shown in Fig. 4.

Heritabilities and repeatabilities

Average h^2 and R, with the corresponding variability, for the three phenotypes analysed are reported in Table 2. The average heritability for FW was mostly moderate, in the range from 0.207 (T1E) to 0.361 (TxE), except for crosses MxR028 (0.422), WxBy (0.468) and BxO (0.783), where high h^2 for FW were estimated. The standard deviation of estimated h^2 ranged from 0.034 (BC2) to 0.392 (TxE). For TA, average h^2 varied between 0.304

(PxF) and 0.832 (BbxNl), with standard deviation in the range 0.005 (PxF) - 0.261 (BC2). Average h^2 of SC ranged from 0.078 (BC2) to 0.861 (BbxNl) (std. dev.: 0.005 (PxF) - 0.269 (TxE)).

Repeatability estimates were on average 51.95% higher than corresponding h^2 estimates: from as little as 0.09% for TA in BbxNl, or 1.99% for FW in BC2, to as much as 273.9% (almost four times) for SC in TxE (from $h^2 = 0.210$ to $R = 0.787$). The variability of repeatability estimates was substantially lower than that of heritability estimates (average coefficient of variation: 12.45% vs 36.05%). Figures 5 and 6 show the boxplots of the 500 h^2 and R estimates per trait and cross.

Predictive ability

Predictive ability (PA, $r(\hat{y}, y)$) was measured in the validation set from a 5-fold cross-validation scheme (Table 2). Each trial was repeated 100 times to assess the variability of PA. Figure 7 reports PA per trait and cross. Moderate values of predictive ability were observed when averaged over crosses: from 0.72 for SC, to 0.65 for TA and 0.60 for FW. When predicting TA the best average performance was achieved in BbxNl (0.83 ± 0.044), the worst
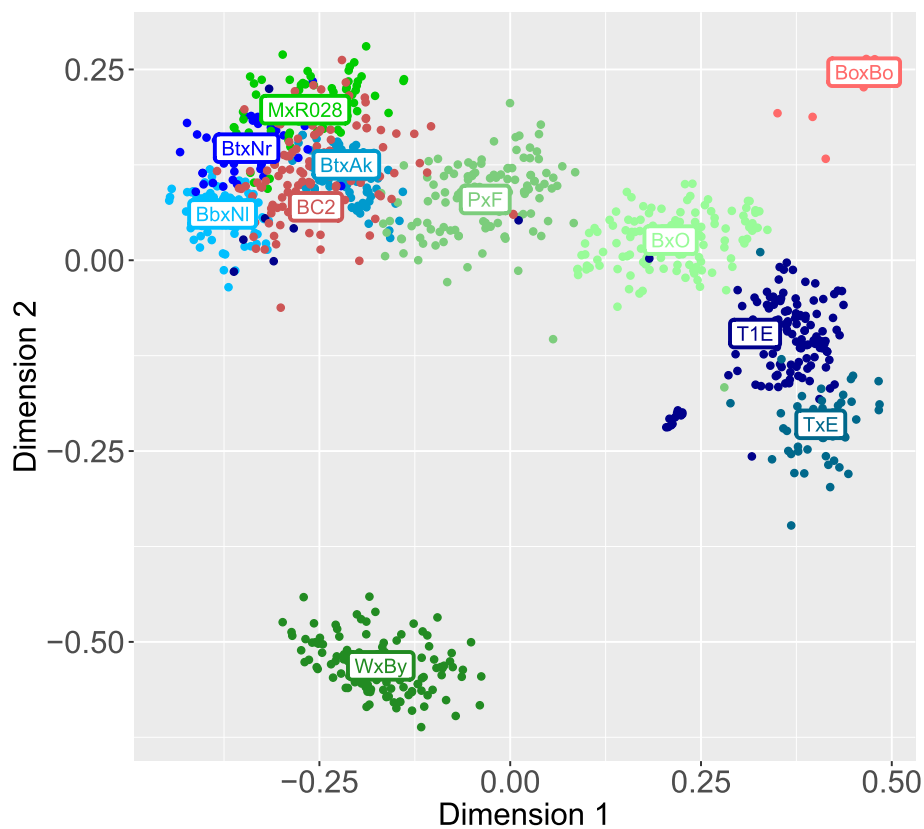

Fig. 3 MDS plot of the matrix of genomic relationships across peach crosses: first two principal coordinates. *Crosses* from France are reported in shades of *red*, from Italy in shades of *green* and from Spain in shades of *blue*. Labels for each progeny are added for clarity

Biscarini *et al. BMC Genomics* (2017) 18:432
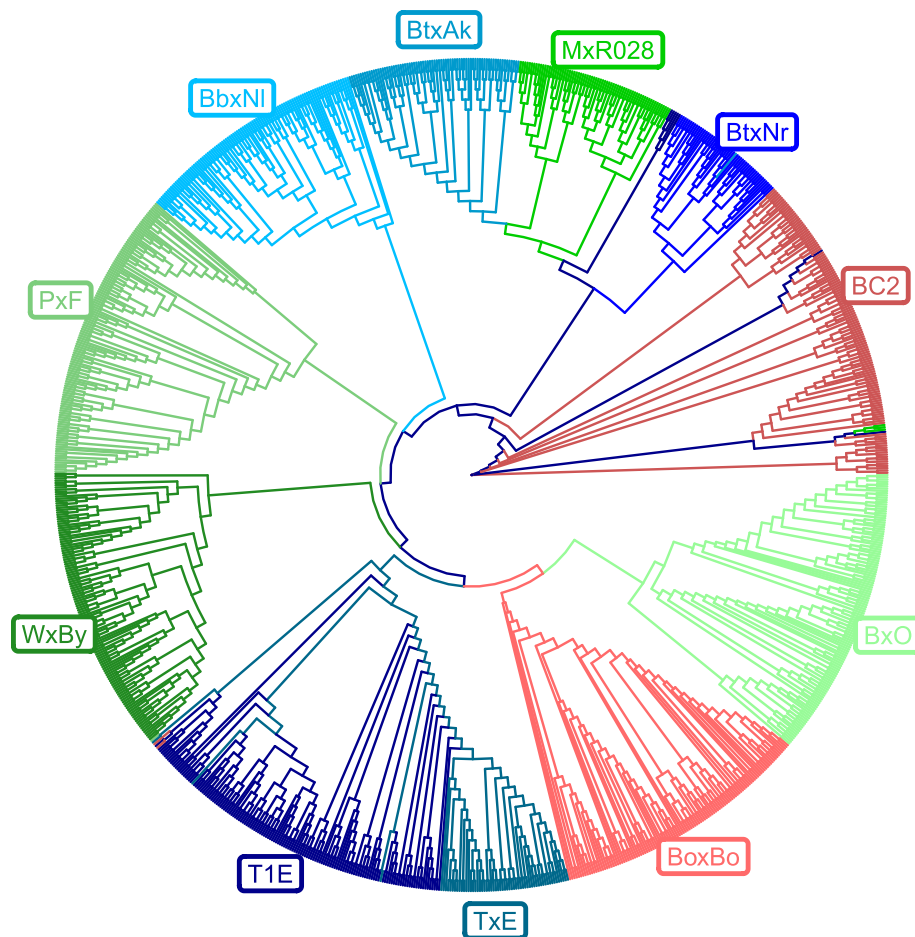
Page 7 of 15

Fig. 4 NJ tree of the 11 peach crosses analysed in this study. *Crosses* from France are reported in shades of *red*, from Italy in shades of *green* and from Spain in shades of *blue*. Labels for each progeny are added for clarity

in PxF (0.5 ± 0.074), with TxE and MxR028 showing high variability of results (standard deviation 0.19 and 0.11, respectively). For FW, the highest predictive ability was estimated in BxO (0.84 ± 0.039) and the lowest in BtxNr (0.39 ± 0.197), with BoxBo displaying by far the largest variability (standard deviation 0.395).

Finally, genomic predictions for SC showed the narrowest range, with average values all falling in the 0.66 (TxE) - 0.78 (BC2) interval. The most variable predictive ability was estimated in TxE (standard deviation 0.16).

Discussion

The joint effort between this paper and the paper by Hernandez-Mora et al. [37] constitutes the first large work to investigate the applicability of genomics-assisted breeding for complex quantitative traits (QTL mapping and genome-enabled predictions) in *P. persica*. The genetic materials and phenotypic records in the two works largely overlap; however, Hernandez-Mora et al. focussed on

QTL detection, and looked at genome-enabled predictions as a collateral result. In this study we provide further insights into genomic predictions in peach trees, considering also the variability of estimates in specific cross-trait combinations. We selected three quantitative continuous traits (fruit weight, sugar content, titratable acidity) based both on their commercial importance and on the availability of complete datasets spanning several crosses and years. The phenotypic information from records in different years has not been pooled, rather used to allow for the effect of permanent environment to be estimated. GBLUP was used for genomic predictions (and estimation of heritability and repeatability), instead of a weighted sum of QTL effects as in FlexQTL®[38]. Additionally, we implemented a repetition protocol to ensure numerical stability in spite of the stochastic variability embedded in cross-validation and Gibbs samplings (BGLR). Finally, the accuracy of imputing missing genotypes in peach was measured in this work.

Biscarini *et al. BMC Genomics*   (2017) 18:432

Page 8 of 15

Table 2 Heritability (h^2), repeatability (R) and predictive ability ($PA: r(\hat{y}, y)$) for fruit weight, sugar content and acidity in the 11 peach populations analysed in this study

Country	Cross	Trait	Avg(h^2)	Sd(h^2)	Avg(R)	Sd(R)	Avg(PA)	Sd(PA)
France	BoxBo	FW	0.22	0.22	0.34	0.28	0.42	0.39
France	BC2	TA	0.63	0.26	0.87	0.03	0.76	0.08
France	BC2	FW	0.24	0.03	0.25	0.03	0.54	0.06
France	BC2	SC	0.08	0.03	0.18	0.03	0.78	0.11
Italy	BxO	TA	0.34	0.07	0.44	0.04	0.52	0.09
Italy	BxO	FW	0.78	0.16	0.83	0.02	0.84	0.04
Italy	BxO	SC	0.63	0.14	0.70	0.03	0.69	0.06
Italy	PxF	TA	0.30	0.00	0.38	0.00	0.50	0.07
Italy	PxF	FW	0.33	0.15	0.45	0.16	0.49	0.24
Italy	PxF	SC	0.53	0.00	0.64	0.00	0.75	0.05
Italy	MxR028	TA	0.39	0.00	0.49	0.00	0.66	0.11
Italy	MxR028	FW	0.42	0.16	0.55	0.09	0.71	0.11
Italy	MxR028	SC	0.46	0.01	0.58	0.00	0.77	0.08
Italy	WxBy	FW	0.47	0.08	0.58	0.06	0.58	0.12
Spain	BbxNl	TA	0.83	0.02	0.83	0.02	0.83	0.04
Spain	BbxNl	FW	0.27	0.20	0.78	0.05	0.63	0.11
Spain	BbxNl	SC	0.86	0.04	0.89	0.03	0.67	0.09
Spain	BtxAk	TA	0.43	0.12	0.46	0.07	0.59	0.10
Spain	BtxAk	FW	0.29	0.21	0.57	0.10	0.70	0.17
Spain	BtxAk	SC	0.62	0.13	0.65	0.03	0.72	0.07
Spain	BtxNr	TA	0.50	0.08	0.65	0.04	0.60	0.11
Spain	BtxNr	FW	0.29	0.18	0.45	0.10	0.39	0.20
Spain	BtxNr	SC	0.57	0.11	0.71	0.04	0.68	0.13
Spain	T1E	TA	0.67	0.08	0.74	0.07	0.79	0.07
Spain	T1E	FW	0.21	0.20	0.54	0.20	0.53	0.29
Spain	T1E	SC	0.41	0.07	0.54	0.06	0.75	0.09
Spain	TxE	TA	0.67	0.13	0.78	0.08	0.58	0.19
Spain	TxE	FW	0.36	0.39	0.86	0.06	0.81	0.13
Spain	TxE	SC	0.21	0.27	0.79	0.08	0.66	0.16

General aspects: heritability, repeatability and estimation model

This study reports a systematic investigation of the applicability of genomic prediction models to key traits for peach fruit quality and marketability. Examples of genomic selection studies in fruit trees include apple [20, 39, 40], Japanese pear [21] and grapevine [41]. A comprehensive review is given by Iwata et al. [42].

Estimated heritability, repeatability and accuracy of genomic predictions varied widely across specific peach crosses and traits. Sample size and phenotypic variability are factors that can affect the absolute value and variability of estimated genetic parameters and genomic predictions. The average sample size (across years) varied dramatically: from 19.6 plants for TA in TxE to 237.5

plants for FW in WxBy. Substantial phenotypic variability was found: the phenotypic coefficient of variation ranged from 16.3 to 48.8% in FW, from 7.2 to 36.9% in SC, and from 10.4 to 73.1% in TA. This reflects the wide range of variability in the peach materials included in the study, and directly influences the estimates of h^2 and R. Additional files 3 and 4 show the coefficient of variation of estimated h^2, R and predictive ability as a function, respectively, of the average sample size and of the phenotypic coefficient of variation. It appears that for larger sample size and phenotypic variance, the estimates of parameters are more reliable (less variable), indicating that these two factors do affect the estimation of heritability, repeatability and the accuracy of genomic predictions.
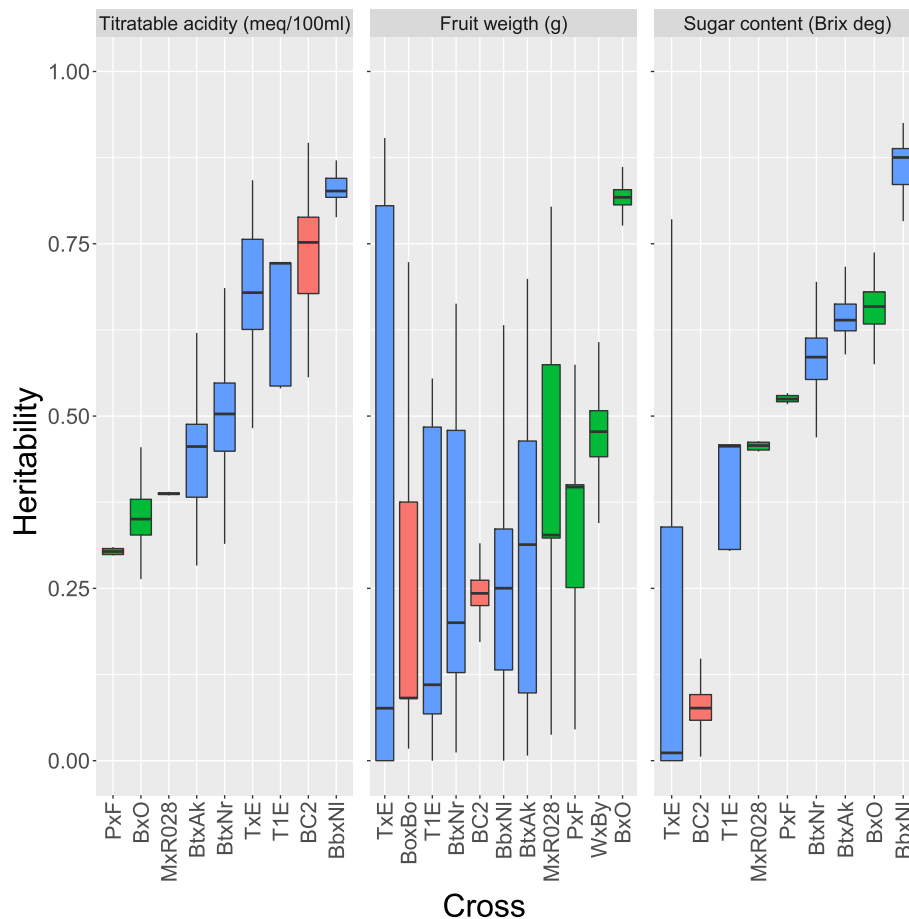
Biscarini *et al. BMC Genomics*   (2017) 18:432

Page 9 of 15

Fig. 5 Boxplots of heritability estimates for acidity, fruit weight and sugar content in 11 peach populations. Results from 5-fold cross-validation repeated 100 times (500 replicates) are presented. *Crosses* from France are reported in *red*, from Italy in *green* and from Spain in *blue*. *Crosses* are ordered by increasing median value

Fruit phenotypes are affected by genetic, ontogenetic (age-related: i.e. consecutive years of growth) and environmental (i.e. climatic years) factors. These can be statistically separated to clarify their contribution to the observed phenotypes (e.g. [43, 44]). From a repeatability model, such as that in Eq. 2, the variance components due to the genetic, permanent and temporary environmental effects are estimated. The permanent environment actually catches the effect of consecutive years of growth, while the temporary environment captures the variability linked to the climatic conditions of specific years. For instance, for sugar content in MxR028, genetics, permanent and temporary environments account for, respectively, 45.6% (h^2), 12.4% ($R - h^2$) and 42% ($1 - R$) of the phenotypic variability. Therefore, the repeatability model may be used as an alternative approach to estimating the genetic, ontogenetic and climatic effects in fruit trees. The less variable the estimates of the genetic parameters, the more reliable the approximations.

Compared to traditional QTL-oriented marker assisted selection, genomic selection is generally thought to perform better for selecting traits controlled by a large number of minor genes, each contributing a small proportion of the total phenotypic variability. The traits used in this study are largely polygenic (FW: [45] SC: [46]) and thus well suited for GBLUP and similar approaches (like SNP/RR-BLUP e.g. [47]), which build on the hypothesis of many small additive allele contributions to the phenotype (i.e. "infinitesimal model" [48]).

We therefore selected a GBLUP framework to apply a repeatability mixed model to the problem of estimating genetic parameters and genomic predictions for fruit weight and quality from SNP genotypes. The model in Eq. 2 was solved either through REML or MCMC, implemented, respectively, in a commercial (ASREML) and an open source (BGLR) software package. This nicely illustrates the difference between statistical model of analysis (GBLUP repeatability model), method of resolution, and specific algorithmic implementation into a software.
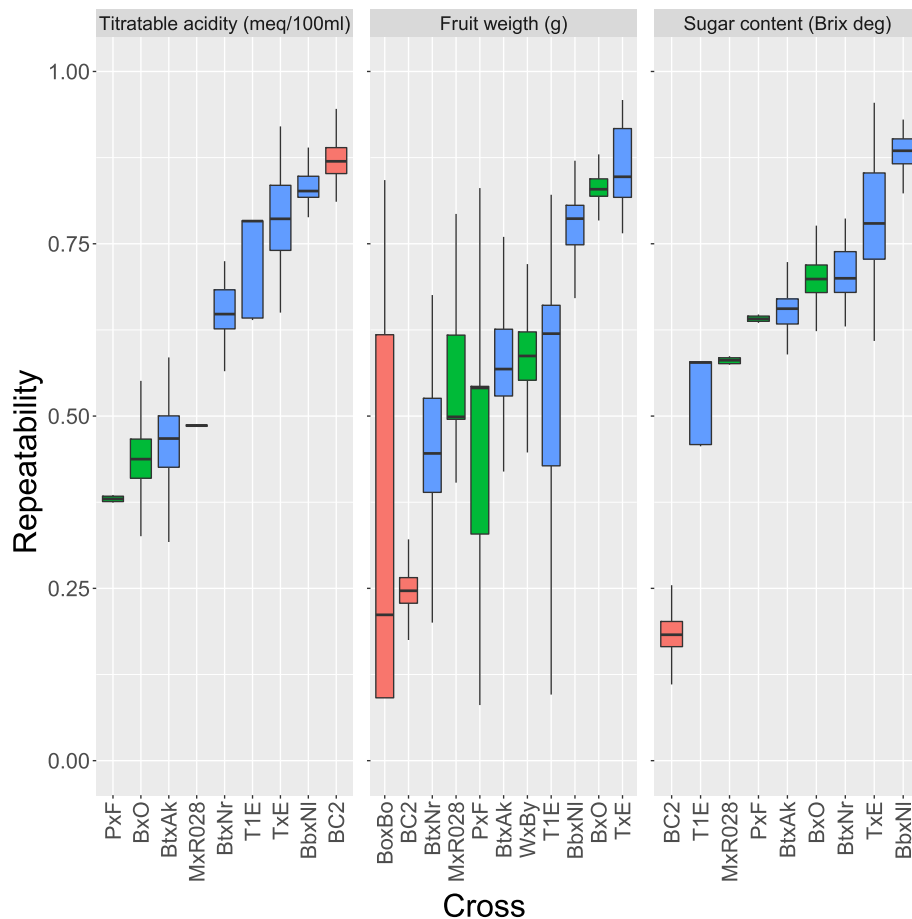
Biscarini *et al. BMC Genomics* (2017) 18:432

Page 10 of 15

Fig. 6 Boxplots of repeatability estimates for acidity, fruit weight and sugar content in 11 peach populations. Results from 5-fold cross-validation repeated 100 times (500 replicates) are presented. *Crosses* from France are reported in *red*, from Italy in *green* and from Spain in *blue*. *Crosses* are ordered by increasing median value

Imputation accuracy

The imputation of missing genotypes has been repeatedly shown to be very accurate: e.g. 95% in humans [49], ∼ 99% in cattle [50], ∼ 98% in rice [28]. Lower imputation accuracy has been observed in other plant species: e.g. 84% in sugar beet [17], ∼ 80% in alfalfa [28]. Imputation errors may have a detrimental effect on the accuracy of genomic predictions (e.g. [51]), and there is therefore interest in assessing imputation accuracy when genotypes are used to predict phenotypes or breeding values. Our results showed that imputation accuracy is quite variable over peach crosses, with a difference of 35.5 percentage points between the best (BC2: 0.96) and the worst (BxO: 0.61) case. The response to increasing missing rates was quite flat in all crosses, with the exception of PxF and BbxNl. Such variable results suggest a strong influence of the genetic background of each population on genotype imputation in peach.

Putting together the average missing rates in the original datasets (Table 1) with the corresponding estimated imputation accuracies (Additional file 2), the amount of imputation errors in the data used for genomic predictions can be estimated in the range 0.88% (BbxNl) - 5.33% (TxE), with an average of 2.71%. Given the low initial missing rates, and the generally good imputation accuracy, there are therefore few residual imputation errors, which are expected to have negligible impact on genomic predictions.

Population structure

Unaccounted population stratification is known to potentially have detrimental effects on genome-wide association studies [52–54] and genomic predictions [41, 55–57]: the association between SNP and phenotype may differ between (sub)populations or be in reverse phase. If possible, it may therefore be advisable, when analysing heterogeneous populations, to account for this in the model (e.g. [58, 59]).

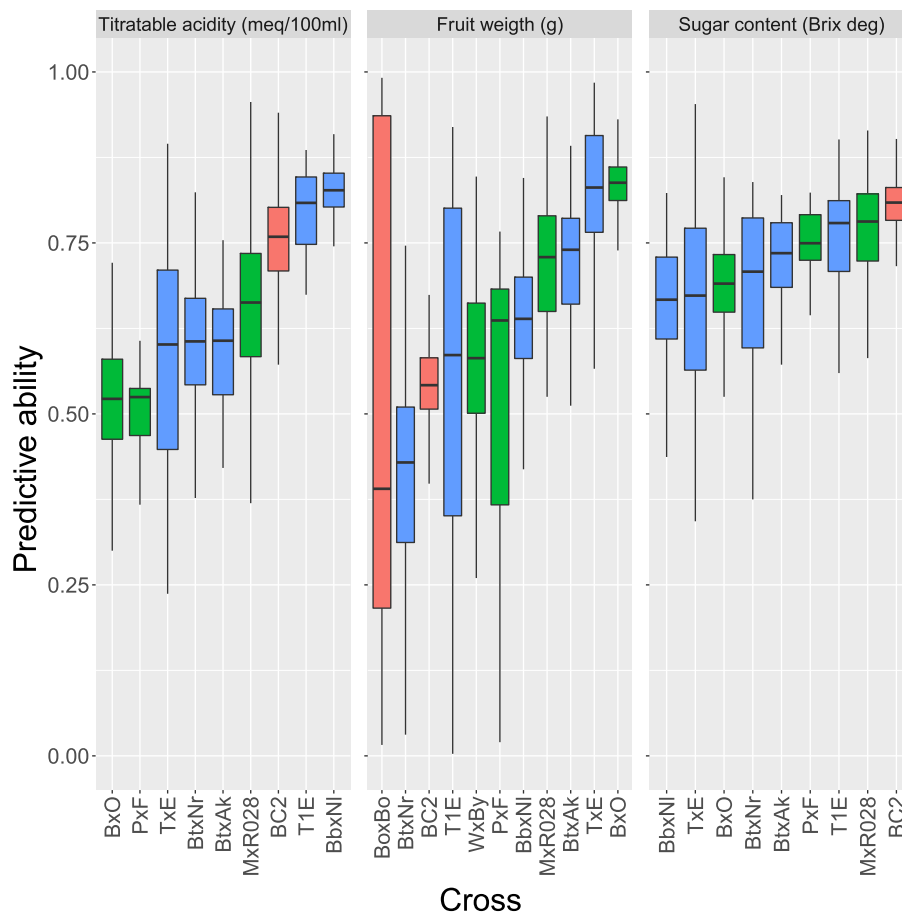In this study, peach crosses were analysed separately. Still, it is interesting to look at population structure,

Biscarini *et al. BMC Genomics*   (2017) 18:432

Page 11 of 15

Fig. 7 Boxplots of the estimated predictive ability for acidity, fruit weight and sugar content in 11 peach populations. Results from 5-fold cross-validation repeated 100 times (500 replicates) are presented. *Crosses* from France are reported in *red*, from Italy in *green* and from Spain in *blue*. *Crosses* are ordered by increasing median value

since this can help interpret the obtained results, and provide guidance for future modeling of genomic predictions in peach populations. The BoxBo resulted in a very compact and isolated cluster, and the relatively limited genetic variability may be related to the comparatively poorer accuracy of genomic predictions in this population. This cross is indeed a self pollination of a partially heterozygous variety. The WxBy cluster is more spread out and prediction accuracy was higher. The separation of this progeny from the other crosses is likely linked with the ornamental NJ Weeping parent —indeed ornamental germplasm is known to have undergone divergent breeding history compared to edible cultivated accessions [60, 61]. The separation of T1E and TxE from the main peach group of progenies from commercial peaches can be attributed to the origin of these populations from almond x peach crosses. The two crosses sharing a parent (T1E and TxE) clustered very closely together. Five crosses (BbxNl, MxR028,

BC2, BtxAk, BtxNr) formed an entangled cluster and higher order dimensions are required to visually separate them.

Accuracy of genomic predictions

The accuracy of genomic predictions for fruit weight, titratable acidity and sugar content was variable across and within crosses, but less so compared to estimates of heritability and repeatability: for FW, the average predictive ability (PA) ranged from 0.39 in BtxNr to 0.84 in BxO; for TA, it ranged between 0.50 in PxF and 0.83 in BbxNl; for SC average PA was in the range 0.66 (TxE) - 0.78 (BC2). The average standard deviation of PA was 0.17, 0.09 and 0.09 for FW, TA and SC respectively. Predictive abilities appear therefore to be more reliable and robust than estimates of variance components due to different sources of variation (additive genetic effects, permanent environment). This is related to the general observation that predictions and inference (e.g. trying to

Biscarini *et al. BMC Genomics* (2017) 18:432

Page 12 of 15

understand the relative influence of genetics and environment on the phenotype) are different problems, and a model may yield good predictions even when the underlying biological mechanisms are poorly understood or estimated (and vice-versa: [62]). In some cases, PA showed very low variability, like in BxO for FW or in BbxNl for TA; in other cases, PA was so variable to become practically unreliable, like FW in BoxBo, where predictive ability went from -0.5 to 0.99. The performance of genomic predictions can be influenced by the size of the analysed dataset, the heritability and repeatability of the trait, and by the phenotypic variability. Additional files 3 and 4 show the coefficient of variation of PA as a function of sample size and phenotypic variability (x-axis). In both figures, the general trend is that the variability of estimates tends to be larger with smaller sample size and smaller phenotypic variability. Additional file 5 shows the coefficient of variability of predictive ability vs the heritability. Again, the larger the heritability, the smaller the variability of predictions, hence their reliability.

Predictive ability is defined as the correlation between the observed phenotype and the phenotype predicted by the model, $r(\hat{y}, y)$. In plant and animal breeding it is often of interest to predict not only the (future) phenotypic value of an individual, but also its unobserved breeding (genetic) value. By dividing predictive abilities by the square root of the heritability of the trait, the accuracy of genomic breeding values (GEBVs) can be estimated [63]:

$$r_{g,\hat{g}} = \frac{r_{y,\hat{y}}}{\sqrt{h^2}} \tag{5}$$

We thus obtained average GEBV accuracies of 0.82, 0.83 and 0.97 for fruit weight, acidity and sugar content, respectively. The higher GEBV accuracy for sugar content reflects the higher average predictive ability (0.72), and the lower variability of estimated heritability (average s.d. 0.088) and repeatability (average s.d. 0.034).

Applications to management and breeding

Genomic selection is having a profound impact on plant breeding. Major drivers behind this success are the possibility of obtaining accurate genomic predictions even without pedigree data, a reference genome or dense marker genotypes, and the higher genetic gains per unit of time that are likely to be achieved (2-3 fold: [64–66]). This is especially true for plant species with a long breeding cycle, since selection candidates can be chosen at a much earlier stage than in traditional breeding programmes. Fruit trees are characterised by a long juvenile phase; in peach breeding programmes, the average generation interval spans 5–7 years, taking into account the length of the juvenile period and repeated years of phenotypic

evaluation, and thus the benefits from shortened cycles through genome-enabled predictions are evident.

The benefits of genomic selection stem not only from accurate genome-enabled predictions and shorter generation intervals, but also from potentially lower phenotyping costs. Collecting phenotypes in fruit trees is costly [67], and restricting phenotypic evaluation to smaller subsets of progenies for shorter times may be beneficial. Coupling genomics, high-throughput phenotyping [68, 69], databases and tools for breeding has the potential of creating effective platforms for genomics-assisted breeding in all plant species (e.g. rice, [70]), particularly in fruit trees, including peach (see Iwata et al. 2016 for a review [42]).

Genome-enabled predictions are mainly used to select breeding candidates in genetic improvement schemes. However, in species with a long life-cycle, accurate genome-enabled predictions may be particularly useful to predict future phenotypes of the plant: e.g. which plants are most likely to repeatedly give a certain production in successive years, which plants to cull, keep, fertilize, which plants are expected to be more resilient to temporary environmental effects (e.g. climatic variations). Additionally, in general breeders/farmers tend to prefer "repeatable/reliable" plants which show little variability in phenotypes from year to year. Resende et al. [71] showed that genomic predictions modelled at early age did not appear to perform well in predicting phenotypes at later ages (6 years). In our study, we modeled multiple records over successive years and obtained accurate genome-enabled predictions of phenotypes in peach. This indicates that there may be value from the application of repeatability GBLUP models in peach breeding.

Conclusions

In this paper, results from a repeatability GBLUP model for fruit weight, sugar content and titratable acidity in peach trees were reported. This is the first work to show the applicability of genomic predictions in *P. persica*. A very diverse set of peach crosses was used, in terms of genetic background of the germplasm, phenotypic variability and, especially, sample size. Still, the obtained results are very promising for the application of genomic selection to peach breeding programmes. The accuracy of imputing missing genotypes was as high as 96%, and the genomic predictive ability was on average 0.65, but could be as high as 0.84 for fruit weight or 0.83 for titratable acidity. The estimated repeatability may prove very useful in the management of the typical long cycles involved in peach productions, since it may indicate which plants bear the potential of being more resilient to temporary fluctuations and give repeatable performances. Additionally, the repeatability model may prove valuable in disentangling genetic, ontogenetic and environmental effects in the analysis of complex traits.

Biscarini *et al. BMC Genomics* (2017) 18:432

Page 13 of 15

All together, the results of this work suggest that the implementation of genomic selection may be very advantageous in *P. persica*, for it can realistically lead to higher genetic gains per unit of time, improved management of the orchard and reduced costs of breeding programs.

Additional files

Additional file 1: Detailed phenotypic data. Phenotypic data per year, cross and trait, comprising descriptive statistics. (CSV 3 kb)

Additional file 2: Imputation accuracy. Statistics on imputation accuracy of missing genotypes, per cross and injected missing levels. (CSV 5 kb)

Additional file 3: Effect of sample size. Figure reporting, for each trait, the coefficient of variation of heritability, predictive ability, and repeatability as functions of sample size. (PDF 6 kb)

Additional file 4: Effect of phenotypic variability. Figure reporting, for each trait, the coefficient of variation of heritability, predictive ability, and repeatability as functions of the coefficient of variation of each phenotipic trait. (PDF 6 kb)

Additional file 5: Coefficient of variation of the predictive ability vs heritability. Figure reporting, for each trait, the coefficient of variation of predictive ability as function of the heritability of the trait in each progeny. Fruit weight in red, Sugar content in green and Titratable acidity in blue. (PDF 5 kb)

Abbreviations

FW: Fruit weight; GBLUP: Genomic best linear unbiased predictions; GEBV: Genomic estimated breeding value; GS: Genomics selection; MAF: Minor allele frequency; PA: Predictive ability; SC: Sugar content; SNP: Single nucleotide polymorphism; TA: Titratable acidity

Availability of data and materials

The data used for this paper are available on-line from the Genome Database for Rosaceae (GDR: https://www.rosaceae.org/publication_datasets).

Authors' contributions

FB and NN performed all statistical analyses, and wrote most of the paper. MB contributed to the development of statistical analyses and writing of the manuscript. PA, MJA, IV, SM, TP, BQT, PL, CdSL, IPC, DB developed genetic materials, carried out phenotyping, and contributed to the writing of the manuscript. AS and LR conceived and coordinated the study and contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1] PTP Science Park, Via Einstein - Loc. Cascina Codazza, Lodi, Italy. [2] Council for Agricultural Research and Economics (CREA) Research Centre for Fodder Crops and Dairy Productions, Lodi, Italy. [3] Wageningen UR Biometris, Wageningen, The Netherlands. [4] IRTA, Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra (Cerdanyola del Vallés), Barcelona, Spain. [5] Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA) — Centro di Ricerca per la Frutticoltura (CREA-FRU), Via di Fioranello 52, Roma, Italy. [6] GAFL, INRA, 84140, Montfavet, France. [7] Università degli Studi di Milano - DiSAA, Via Celoria 2, Milano, Italy. [8] IBBA-CNR, Via Edoardo Bassini, 15, 20133 Milan, Italy. [9] Institute of Nutrition and Food Technology - INTA, Universidad de Chile, Av El Líbano 5524, Santiago, Chile. [10] Present Address: Hendrix Genetics Research, Technology & Services B.V., P.O. Box 114, 5830AC Boxmeer NL, The Netherlands.

References

1. Faust M, Timon B, et al. Origin and dissemination of peach. Hort Rev. 1995;17:331–79.
2. Abbott A, Georgi L, Yvergniaux D, Inigo M, Sosinski B, Wang Y, Blenda A, Reighard G. Peach: the model genome for rosaceae. Acta Horticulturae. 2002;1:145–56.
3. FAOSTAT, FAO. http://faostat.fao.org. Accessed 25 Jan 2017.
4. Luby JJ, Shaw DV. Does marker-assisted selection make dollars and sense in a fruit breeding program? HortScience. 2001;36(5):872–9.
5. Monet R, Bassi D. Classical genetics and breeding. The Peach. Botany, Production and Uses. Wallingford: CAB International; 2008, pp. 61–84.
6. Lin Z, Hayes B, Daetwyler H. Genomic selection in crops, trees and forages: a review. Crop Pasture Sci. 2014;65(11):1177–91.
7. Hayes B, Bowman P, Chamberlain A, Goddard M. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci. 2009;92(2):433–43.
8. Nat. Biotechnol E. The genome-assisted barnyard. Nat Biotechnol. 2009;27:487.
9. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. PloS ONE. 2011;6(5):19379.
10. Martínez-García PJ, Parfitt DE, Ogundiwin EA, Fass J, Chan HM, Ahmad R, Lurie S, Dandekar A, Gradziel TM, Crisosto CH. High density snp mapping and qtl analysis for fruit quality characteristics in peach (prunus persica l). Tree Genet Genomes. 2013;9(1):19–36.
11. Nicolazzi E, Biffani S, Biscarini F, Orozco ter Wengel P, Caprera A, Nazzicari N, Stella A. Software solutions for the livestock genomics snp array revolution. Animal Genet. 2015;46(4):343–53.
12. Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, et al. The high-quality draft genome of peach (prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet. 2013;45(5):487–94.
13. Verde I, Jenkins J, Dondini L, Micali S, Pagliarani G, Vendramin E, Paris R, Aramini V, Gazza L, Rossini L, et al. The peach v2. 0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. BMC Genomics. 2017;18(1):225.
14. Verde I, Bassil N, Scalabrin S, Gilmore B, Lawley CT, Gasic K, Micheletti D, Rosyara UR, Cattonaro F, Vendramin E, et al. Development and evaluation of a 9k snp array for peach by internationally coordinated snp detection and validation in breeding germplasm. PLoS ONE. 2012;7(4):35668.
15. Annicchiarico P, Nazzicari N, Li X, Wei Y, Pecetti L, Brummer EC. Accuracy of genomic selection for alfalfa biomass yield in different reference populations. BMC Genomics. 2015;16(1):1020.
16. Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y. Genomic selection in sugar beet breeding populations. BMC Genetics. 2013;14(1):85.

Biscarini *et al. BMC Genomics*   (2017) 18:432

Page 14 of 15

17. Biscarini F, Stevanato P, Broccanello C, Stella A, Saccomani M. Genome-enabled predictions for binomial traits in sugar beet populations. BMC Genetics. 2014;15(1):87.

18. Resende MF, Muñoz P, Resende MD, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M. Accuracy of genomic selection methods in a standard data set of loblolly pine (pinus taeda l). Genetics. 2012;190(4):1503–10.

19. Resende MD, Resende MF, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, et al. Genomic selection for growth and wood quality in eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol. 2012;194(1):116–28.

20. Kumar S, Chagné D, Bink MC, Volz RK, Whitworth C, Carlisle C. Genomic selection for fruit quality traits in apple (malus× domestica borkh). PloS ONE. 2012;7(5):36674.

21. Iwata H, Hayashi T, Terakami S, Takada N, Sawamura Y, Yamamoto T. Potential assessment of genome-wide association study and genomic selection in japanese pear pyrus pyrifolia. Breed Sci. 2013;63(1):125–40.

22. Guo G, Zhao F, Wang Y, Zhang Y, Du L, Su G. Comparison of single-trait and multiple-trait genomic prediction models. BMC Genetics. 2014;15(1):1.

23. Mrode RA. Linear models for the prediction of animal breeding values. Cambridge: Cabi; 2014.

24. Wardyn BM, Edwards JW, Lamkey KR. Inbred-progeny selection is predicted to be inferior to half-sib selection for three maize populations. Crop Sci. 2009;49(2):443–50.

25. Cavalcanti JJV, de Resende MDV, Crisóstomo JR, de Moura Barros L, de Paiva JR. Genetic control of quantitative traits and hybrid breeding strategies for cashew improvement. Crop Breed Appl Technol. 2007;7(2):186.

26. Micheletti D, Dettori MT, Micali S, Aramini V, Pacheco I, Linge CDS, Foschi S, Banchi E, Barreneche T, Quilot-Turion B, et al. Whole-genome analysis of diversity and snp-major gene association in peach germplasm. PloS ONE. 2015;10(9):0136803.

27. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007;81(5):1084–1097.

28. Nazzicari N, Biscarini F, Cozzi P, Brummer EC, Annicchiarico P. Marker imputation efficiency for genotyping-by-sequencing data in rice (oryza sativa) and alfalfa (medicago sativa). Mol Breeding. 2016;36(6):1–16.

29. Astle W, Balding DJ. Population structure and cryptic relatedness in genetic association studies. Stat Sci. 2009;24(4):451–71.

30. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.

31. Goddard M, Hayes B, Meuwissen T. Using the genomic relationship matrix to predict the accuracy of genomic selection. J Anim Breeding Genet. 2011;128(6):409–21.

32. Patterson HD, Thompson R. Recovery of inter-block information when block sizes are unequal. Biometrika. 1971;58(3):545–54.

33. Casella G, George EI. Explaining the gibbs sampler. Am Stat. 1992;46(3):167–74.

34. Gilmour AR, Gogel BJ, Cullis BR, Thompson R, Butler D. Asreml user guide release 3.0. Hemel Hempstead: VSN International Ltd; 2009.

35. Pérez P, de Los Campos G. Genome-wide regression & prediction with the bglr statistical package. Genetics. 2014;206(1):114.

36. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2016. R Foundation for Statistical Computing. https://www.R-project.org/.

37. Hernández Mora JR, Micheletti D, Bink M, Van de Weg E, Cantín C, Nazzicari N, Caprera A, Detori MT, Micali S, Banchi E, Campoy JA, Dirlewanger E, Lambert P, Pascal T, Troggio M, Bassi D, Rossini L, Verde I, Quilot-Turion B, Laurens F, Arús P, Aranzana MJ. Integrated QTL detection for key breeding traits in multiple peach progenies. BMC Genomics. 2017. doi:10.1186/s12864-017-3783-6.

38. Bink M, Uimari P, Sillanpää M, Janss L, Jansen R. Multiple qtl mapping in related plant populations via a pedigree-analysis approach. Theor Appl Genet. 2002;104(5):751–62.

39. Kumar S, Bink MC, Volz RK, Bus VG, Chagné D. Towards genomic selection in apple (malus× domestica borkh) breeding programmes: prospects, challenges and strategies. Tree Genet Genomes. 2012;8(1):1–14.

40. Muranty H, Troggio M, Sadok IB, Al Rifaï M, Auwerkerken A, Banchi E, Velasco R, Stevanato P, Van De Weg WE, Di Guardo M, et al. Accuracy and responses of genomic selection on key traits in apple breeding. Hortic Res. 2015;2:15060.

41. Fodor A, Segura V, Denis M, Neuenschwander S, Fournier-Level A, Chatelet P, Homa FAA, Lacombe T, This P, Le Cunff L. Genome-wide prediction methods in highly diverse and heterozygous species: proof-of-concept through simulation in grapevine. PloS ONE. 2014;9(11):110436.

42. Iwata H, Minamikawa MF, Kajiya-Kanegae H, Ishimori M, Hayashi T. Genomics-assisted breeding in fruit trees. Breed Sci. 2016;66(1):100.

43. Loughin T. Improved experimental design and analysis for long-term experiments. Crop Sci. 2006;46:2492–502.

44. Segura V, Durel C, Costes E. Dissecting apple tree architecture into genetic, ontogenetic and environmental effects: Qtl mapping. Tree Genet Genomes. 2009;5(1):165–79.

45. da Silva Linge C, Bassi D, Bianco L, Pacheco I, Pirona R, Rossini L. Genetic dissection of fruit weight and size in an f2 peach (prunus persica (l.) batsch) progeny. Mol Breeding. 2015;35(2):1–19.

46. Cirilli M, Bassi D, Ciacciulli A. Sugars in peach fruit: a breeding perspective. Hortic Res. 2016;3:15067.

47. Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G. The impact of genetic relationship information on genomic breeding values in german holstein cattle. Genetic Sel Evol. 2010;42(1):1.

48. Meuwissen T, Goddard M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics. 2010;185(2):623–31.

49. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet. 2006;78(4):629–44.

50. Weigel K, de Los Campos G, Vazquez A, Rosa G, Gianola D, Van Tassell C. Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in jersey cattle. J Dairy Sci. 2010;93(11):5423–435.

51. Rutkoski JE, Poland J, Jannink JL, Sorrells ME. Imputation of unordered markers and the impact on genomic selection accuracy. G3: Genes Genomes Genet. 2013;3(3):427–39.

52. Lander ES, Schork NJ, et al. Genetic dissection of complex traits. Science-New York then Washington. 1994;265:2037–47.

53. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. Nat Genet. 2004;36(5):512–7.

54. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet. 2010;11(7):459–63.

55. Zhong S, Dekkers JC, Fernando RL, Jannink JL. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics. 2009;182(1):355–64.

56. Janss L, de Los Campos G, Sheehan N, Sorensen D. Inferences from genomic models in stratified populations. Genetics. 2012;192(2):693–704.

57. Guo Z, Tucker DM, Basten CJ, Gandhi H, Ersoz E, Guo B, Xu Z, Wang D, Gay G. The impact of population structure on genomic prediction in stratified populations. Theor Appl Genet. 2014;127(3):749–62.

58. Biscarini F, Bovenhuis H, Van Der Poel J, Rodenburg T, Jungerius A, Van Arendonk J. Across-line snp association study for direct and associative effects on feather damage in laying hens. Behav Genet. 2010;40(5):715–27.

59. Biscarini F, Bovenhuis H, Van Arendonk J, Parmentier H, Jungerius A, Van Der Poel J. Across-line snp association study of innate and adaptive immune response in laying hens. Anim Genet. 2010;41(1):26–38.

60. Cao K, Zheng Z, Wang L, Liu X, Zhu G, Fang W, Cheng S, Zeng P, Chen C, Wang X, et al. Comparative population genomics reveals the domestication history of the peach, prunus persica, and human influences on perennial fruit crops. Genome Biol. 2014;15(7):1.

61. Akagi T, Hanada T, Yaegaki H, Gradziel TM, Tao R. Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication. DNA Res. 2016;23(3):271–82.

62. de los Campos G, Sorensen D, Gianola D. Genomic heritability: what is it? PLoS Genet. 2015;11(5):1005048.

63. Legarra A, Robert-Granié C, Manfredi E, Elsen JM. Performance of genomic selection in mice. Genetics. 2008;180(1):611–8.

Biscarini *et al. BMC Genomics*   (2017) 18:432

Page 15 of 15

64. Grattapaglia D, Resende MD. Genomic selection in forest tree breeding. Tree Genet Genomes. 2011;7(2):241–55.

65. Iwata H, Hayashi T, Tsumura Y. Prospects for genomic selection in conifer breeding: a simulation study of cryptomeria japonica. Tree Genet Genomes. 2011;7(4):747–58.

66. Denis M, Bouvet JM. Efficiency of genomic selection with models including dominance effect in the context of eucalyptus breeding. Tree Genet Genomes. 2013;9(1):37–51.

67. Myles S. Improving fruit and wine: what does genomics have to offer? Trends Genet. 2013;29(4):190–6.

68. Poland J. Breeding-assisted genomics. Curr Opin Plant Biol. 2015;24: 119–24.

69. Araus JL, Cairns JE. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. 2014;19(1):52–61.

70. Biscarini F, Cozzi P, Casella L, Riccardi P, Vattari A, Orasen G, Perrini R, Tacconi G, Tondelli A, Biselli C, et al. Genome-wide association study for traits related to plant and grain morphology, and root architecture in temperate rice accessions. PloS ONE. 2016;11(5):0155425.

71. Resende M, Munoz P, Acosta J, Peter G, Davis J, Grattapaglia D, Resende M, Kirst M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol. 2012;193(3):617–24.